Electronic Supplementary Information

Recombinant perlucin derivatives influence the nucleation of calcium carbonate

E. Weber, I. M. Weiss, H. Cölfen and M. Kellermeier

Supplementary Figures

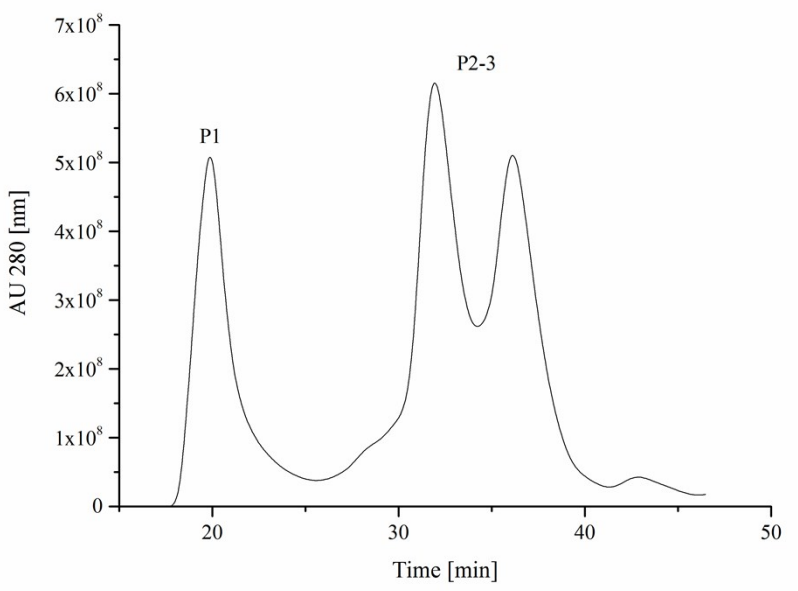

Fig. S1 Representative SEC elution profile of GFP-perlucin, showing that the protein elutes in three different peak fractions (P1, P2 and P3).

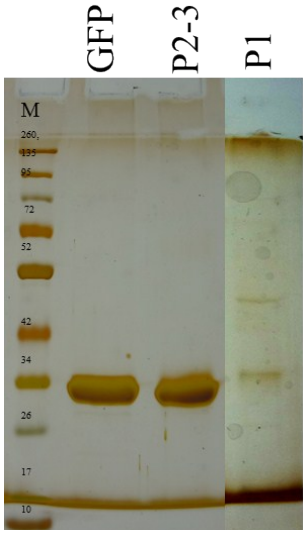

Fig. S2 SDS-PAGE analysis of GFP and the GFP-perlucin derivatives P2-3 and P1 after their purification by size-exclusion chromatography.

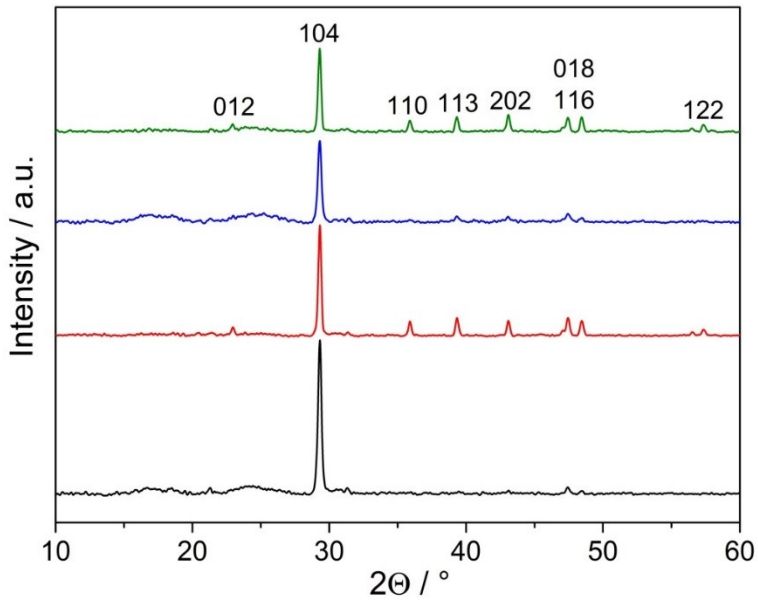

Fig. S3 Powder XRD patterns recorded from precipitates isolated at the end of titration experiments without protein (black) and in the presence of 0.01 mg/mL GFP (red), P1 (blue), and P2-3 (green). All occurring reflections can be assigned to calcite planes as indicated.

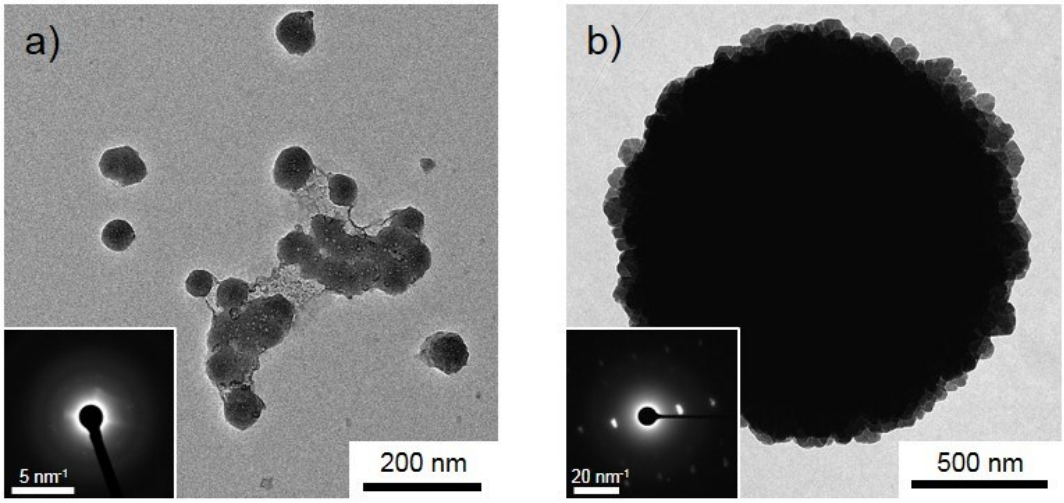

Fig. S4 TEM images and corresponding electron diffraction (ED) patterns (insets) of a) amorphous nanoparticles formed in the absence of proteins, and b) crystalline aggregates grown under the influence of 0.01 mg/mL P1. Note that the particles were isolated from separate titration experiments that were stopped soon after the maximum in the free amount of calcium was reached and nucleation took place. The results thus show that crystalline particles were present already during the very early post-nucleation stages in the protein-containing sample, whereas only amorphous species were observed in the reference at the same stage.

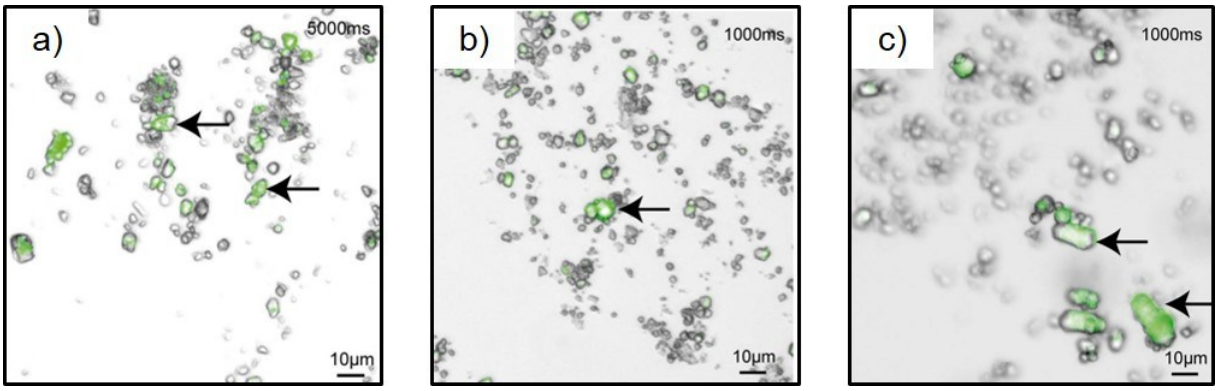

Fig. S5 Overlay of fluorescence and bright-field micrographs of calcite particles isolated at the end of titration experiments in the presence of 0.01 mg/mL of a) GFP, b) P1 and c) P2-P3. Fluorescence of GFP at the surface of the crystals (especially the elongated ones; indicated by arrows) confirms protein adsorption. The exposure times were 5000 ms in (a) and 1000 ms in (b) and (c).

Supplementary Discussion: Protein Sequences

As described in a previous study,[S1] the protein fraction P1 (GFP-perlucin "full-length" variant) corresponds to a mixture of perlucin derivatives consisting of 395-410 amino acids in total, of which 258 (N-terminal) belong to the GFP part. The longest GFP-perlucin version detected in this fraction contains 13 acidic and 13 basic amino acids in its central domain (interspaced by multiple aromatic and cysteine residues) and is terminated by two repetitive "HANLQQRD" sequence motifs,[S2,S3] which likely have a major impact on protein solubility and agglomeration (note that full-length recombinant perlucin only becomes soluble after attachment to GFP). The truncated variants in fraction P2-3 lack most of the perlucin sequence, including the terminal repeats with the "RD" motif, and differ from GFP by about 7-11 amino acids (giving a maximum length of 265-269 amino acids).[S1] An interesting common feature of the proteins in fractions P1 and P2-3 is the occurrence of "RR" pairs (multiple such pairs in P1, one in P2-3), which are absent in GFP alone. We note that all perlucin derivatives used in the present work were equal to those characterised biochemically in the mentioned previous study.[S1]

The presence of "RD" in the repeat motif of P1 suggests that the influence of these proteins on mineralisation and their aggregation behaviour depend on pH. At high pH, the arginine units become neutral by deprotonation, while aspartate retains its negative charge, rendering ionic interactions between the repeat domains (and/or perlucin's central domain) less likely, although calcium ions from the mineralisation process may act as an intra- and/or intermolecular cationic bridge (along with disulfide bridges formed by the cysteine residues in the central domain). In the present work, the pH of the solutions was kept constant at 9 (in contrast to previous studies,[S1-S3] which dealt with lower pH values) and hence the "RD" units are partly zwitterionic and partly negatively charged, possibly allowing them to act as a "switch" in the interactions between the proteins and mineral precursors. Even in the presence of other proteins lacking such domains, these specific functional groups and their dynamic charge-switching effect could take over a dominant role in protein-ion and/or protein-mineral

interactions. In fraction P2-3, the central perlucin domain and the repeat units are missing, suggesting that these parts of the P1 proteins account for the observed activating effect on $CaCO_3$ crystallisation. GFP itself is rich in acidic residues, which likely cause its retarding influence. Finally, the proteins in fraction P2-3 differ from GFP primarily in terms of the striking "RR" pair (partly neutral and partly positively charged at pH 9), which may provide additional inhibiting power in the fusion protein compared to GFP alone. Alternatively, the N-terminal part of perlucin (present in P2-3, absent in GFP) could also cause some sterical hindrance that may alter the protein-ion or protein-mineral interactions – although all of these arguments are clearly speculative and await confirmation by future studies.

Experimental Details

Cloning and protein expression. Recombinant histidine-tagged GFP and histidine-tagged GFP-perlucin were expressed in *E.coli* host systems as described previously.[S1] Cloning was achieved using pQE expression systems (Qiagen, Hilden, Germany) based on the pQE31-GFP-CBP vector.[S4] Cells from single colonies were cultivated for 14 hours in 5 mL Luria broth (LB) medium at 37°C, 180 rpm for inoculation of 100 mL LB medium. Cells were grown to O.D.600 ~ 0.5-0.7 prior to 0.5 mM IPTG induction (Fermentas, St. Leon-Rot, Germany), and continued to grow at 22°C for 8 hours and at 18°C for 16 hours at 180 rpm. Cells were harvested by centrifugation at 3,200 g for 20 minutes at 4°C shock frozen in liquid nitrogen. Protein was purified according to the manufacturer's instructions using 100 µL (50%) Ni^{2+}-NTA matrix (Ni^{2+}-NTA Agarose, Qiagen, Hilden, Germany) per 100 mL in 50 mL reaction tubes. GFP was purified from the respective host cells, harvested and natively extracted according to standard procedures as described elsewhere.[S5]

Size exclusion chromatography. One step eluate of affinity-purified GFP and GFP-perlucin was prepared as described in previous work.[S1] Proteins were concentrated in centrifugal membrane devices (3 kDa MWCO, Pall, Ann Arbor, USA) up to ~6 mg/mL. Concentrated eluates were subjected to size exclusion chromatography (SEC) using an ÄKTA Purifier system equipped with a Superose 12 10/300 GL column (GE Healthcare, Munich, Germany). Peak fractions were collected in elution buffer (300 mM NaCl, 50 mM NaH_2PO_4, 500 mM imidazole buffer pH 8.0) at a constant flow rate of 0.4 mL/min and a detection wavelength of 280 nm. Quantification of protein in the SEC peak fractions was achieved by means of a colorimetric Bradford assay (Sigma-Aldrich, Munich, Germany) using a standard calibration curve.

Protein characterisation. The purified proteins were characterised by standard reducing SDS-PAGE analysis. Protein bands were stained as described elsewhere.[S6] A Spectra™ Multicolor broad range ladder (#SM1841, Fermentas, St. Leon-Rot, Germany) was used as a molecular weight marker.

Potentiometric titrations. All titration experiments were performed using a commercial system from Metrohm (Filderstadt, Germany) and procedures documented in detail in the literature.[S7] The setup consisted of a Titrando 809 titration device with two Dosino 807 dosing units and sensors for monitoring the pH (flat-membrane glass electrode, Metrohm No. 6.0256.100) and calcium potential (polymer-based ISE, Metrohm No. 6.0508.110) in the analysed solutions. Carbonate buffers were prepared by mixing 10 mM solutions of sodium carbonate (Aldrich, reagent grade) and sodium bicarbonate (Riedel de-Haën, reagent grade) in appropriate ratios to give a pH of 9.00. Proteins were transferred from their original elution buffer by concentration in centrifugal membrane devices (3 kDa MWCO) and subsequent dilution in carbonate buffer, a process that was repeated several times to ensure complete removal of any other salts. Eventually, the protein concentration was adjusted to 0.01 mg/mL. For measurement, 15 mL of carbonate buffer (with or without protein) was filled into a beaker, and 0.01 M $CaCl_2$ solution (dilution of 1 M volumetric standard from Fluka) was titrated at a rate of 0.01 mL/min while automated counter-titration of 0.01 M NaOH (Alfa Aesar, standard solution) kept the pH of the buffer constant at 9.00 ± 0.05. Dosing control and data readout was done by customised software (Tiamo, v2.2). The pH electrode was calibrated weekly with buffers from Mettler-Toledo (pH 4, 7 and 9). Calibration of the Ca^{2+}-ISE was achieved by titration of 0.01 M $CaCl_2$ into water under the same conditions. Data were evaluated under the assumption of ideal solutions, as described elsewhere.[S7,S8] All solutions were prepared with water of Milli-Q quality. During titrations, the beakers were continuously flushed with a gentle stream of nitrogen to avoid CO_2 uptake from the atmosphere.

Characterisation of mineral particles. In order to investigate the morphology and structure of particles formed in the titration assays, solutions were passed through membrane filters (Whatman, 200 nm) at the end of the experiments to isolate any solid material, which was briefly washed with water and subsequently dried in air. For scanning electron microscopy (SEM), the isolated powders were applied to aluminium stubs by using double-sided adhesive tape. SEM analyses were carried out on a Hitachi TM-3000 tabletop microscope operated at 15 kV in backscattered electron detection mode. X-ray diffraction (XRD) patterns of the particles were acquired on a Bruker AXS D8 Advance diffractometer over a 2θ range of 10-60° with a scan time of 1 h. Fluorescence microscopy was performed using an inverted microscope from Zeiss (Observer Z1) equipped with A-Plan 10x/0.25Ph1, LD Plan Neofluar 20x/0.4 Korr Ph2 objectives and a Colibri fluorescence system. For transmission electron microscopy (TEM), aliquots of titration solutions were directly applied to carbon-coated copper grids (Quantifoil, 400 mesh) and blotted with the aid of a filter paper to generate a thin film that was then dried in air. TEM images and ED patterns were acquired on a JEOL JEM-2200FS microscope at an acceleration voltage of 240 kV.

References

S1. E. Weber, C. Guth and I. M. Weiss, *Plos One*, 2012, **7**, e56653.

S2. K. Mann, I. M. Weiss, S. Andre, H. J. Gabius and M. Fritz, *Eur. J. Biochem.*, 2000, **267**, 5257-5264.

S3. T. Dodenhof, F. Dietz, S. Franken, I. Grunwald and S. Kelm, *Plos One*, 2014, **9**, e97126.

S4. I. M. Weiss and V. Schönitzer, *J. Struct. Biol.*, 2006, **153**, 264-277.

S5. *The QIAexpressionist™ – A handbook for high-level expression and purification of 6xHis-tagged proteins*, Qiagen (Hilden, Germany), 5th edition, 2003.

S6. H. Blum, H. Beier and H. J. Gross, *Electrophoresis*, 1987, **8**, 93–99.

S7. M. Kellermeier, H. Cölfen and D. Gebauer, *Methods in Enzymology*, 2013, **532**, 45-69.

S8. M. Kellermeier, A. Picker, A. Kempter, H. Cölfen and D. Gebauer, *Adv. Mater.*, 2014, **26**, 752-757.